

<https://helda.helsinki.fi>

Lexical bundles in Early Modern and Present-day English Acts of Parliament

Lehto, Anu

John Benjamins

2018

Lehto , A 2018 , Lexical bundles in Early Modern and Present-day English Acts of Parliament . in J Kopaczyk & J Tyrkkö (eds) , Applications of Pattern-driven Methods in Corpus Linguistics . Studies in Corpus Linguistics , no. 82 , John Benjamins , Amsterdam , pp. 159-186 . <https://doi.org/10.1075/scl.82.07leh>

<http://hdl.handle.net/10138/326150>

<https://doi.org/10.1075/scl.82.07leh>

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Lexical bundles in Early Modern and Present-day English Acts of Parliament

Anu Lehto

University of Helsinki

This chapter analyses three-word sequences in Early Modern and Present-day English legal writing by defining their grammatical and functional distribution in Acts of Parliament. The method follows a corpus-driven approach: the lexical bundles are retrieved automatically from the corpus using frequency as the criterion. The study indicates that lexical bundles in acts extend to the textual level and reveals consistent word combinations on the level of the lexis. The study illustrates that the acts are established as a genre, and the overall distribution of both grammatical types and functions of bundles is rather similar in all the analysed periods. Nevertheless, textual organisation is more important in contemporary acts and textual links further become more specific, although early modern bundles already show textual patterning. Noun phrase and prepositional phrases also increase in contemporary acts, indicating a change to nominal writing conventions.

Keywords: lexical bundles; legal writing; textual organisation; Early Modern English; Present-day English

1. Introduction

This chapter explores three-word bundles in Early Modern and Present-day English Acts of Parliament. Acts are valuable data for the analysis of lexical bundles, as the genre is generally considered to be formulaic (e.g. Danet 1980; Breeze 2013). Legal writing has additionally faced certain diachronic changes since the layout of the acts, for instance, has become more regular. The study includes material from historical and present-day eras to trace possible formulaic wordings and diachronic changes in the writing conventions. The acts can be expected to be formulaic in their wording, but the study uncovers diachronic developments as well: the grammatical and functional distribution of the bundles indicates that the acts developed toward nominal writing conventions and that textual patterning became more important.

Earlier studies on word combinations vary in their approaches. Moon (1998), for instance, retrieves known and pre-established idioms for survey, while Biber, Conrad and Cortes (2004) and Hyland (2008) rely on a corpus-driven method. The latter two studies recognise lexical bundles based on their frequency and distribution over a set number of corpus texts (see Section 2). My study is based solely on a corpus-driven approach, as the bundles were retrieved automatically from corpora instead of utilising pre-established lists (see Tognini-Bonelli 2001: 65ff.; McEnery, Xiao & Tono 2006: 8; cf. Meyer 2015). In a strict sense, a corpus-driven methodology indicates that the theoretical framework is established on the basis of the findings, but the corpus-driven and corpus-based approaches are often used at the same time (see Meyer 2015). Consequently, my study employs the corpus-driven method in detecting bundles; however, existing categorisations of grammatical types and functions of bundles are used in the analysis, and these taxonomies are then further altered to fit the current findings.

The lexical bundle method alone does not sufficiently capture repetition and formulaic structures in legal writing since much of that repetition is based on formulaic text structures and layout. Although the approach is corpus-driven, the investigation emphasises the context of the bundles and the texts. In general, the corpus-driven method reveals lexical strings that show recurrent word combinations and, for example, fragments of dependent clauses. Contemporary acts use a specific layout that is based on tabulation and numbering, and many of the bundles contain letters and other fragments that are part of the textual structure. The layout used in contemporary acts is shown below in (1):

- (1) (3) The Treasury may—
 (a) issue money out of the Consolidated Fund, and
 (b) apply it in the year ending with 31 March 2016 for expenditure authorised by
 Parliament, up to the amount of £219,606,746,000.
 (Supply and Appropriation (Anticipation and Adjustments) Act 2015 c. 10, section 1)

The beginning of the sentence, *The Treasury may*, is numbered, and the two permissions that continue the sentence are placed on separate lines beginning with points *a* and *b*. This layout is not yet found in the historical acts, but the textual structure is visible in repeated *that*-clause fragments and in bundles based on the enacting formula (*and be it*). My study illustrates that these bundles indicate shifts in the structure of the text at the same time, and they are further emphasised by paragraph changes or paragraph marks, see (3) below. The bundles thus coincide with the superstructure, which is a formulaic scheme of a particular genre (see van Dijk 1985: 69; Cortes 2013). The study further establishes functional categories for bundles that show textual organisation, as these bundles contain not only repeated word combinations but also display repetition on the level of the text.

The material in the study are acts from the *Corpus of Early Modern English Statutes* (1491–1689) and from a corpus of contemporary UK Acts of Parliament from 2015 (see Section 3). The study assesses three-word lexical bundles by their grammatical structures and functions, and compares the distribution of lexical bundles in both the historical and the contemporary acts. Additionally, the early modern data is scrutinised diachronically in order to illustrate that some of the linguistic changes had taken place already during this period. The study shows that textual organisation has become more established in the contemporary data and the textual links are more specific. The main grammatical categories and functional types, nevertheless, remain rather similar in legal writing from the Early Modern era onwards. The repeated sequences are mainly noun and prepositional phrases in both periods, and the referential bundles prevail. The bundles still display some grammatical changes, for example, noun and prepositional phrase fragments become proportionally more frequent in the contemporary acts, while dependent clause fragments decline, reflecting a change toward literacy and a more nominal style of writing.

2. The definition of lexical bundles

Lexical bundles are repeated word strings of three or more words, and they appear more often than random word combinations in texts (e.g. Biber et al. 1999; Hyland 2008). Bundles are recognised solely on their frequency and distribution (cf. Moon 1998). In earlier studies, thresholds for frequencies range mainly from 10 to 40 instances in a million words, and dispersion varies from 5 to 10. My study sets the frequencies to more than 50 and 70 instances

in a million words, as the three-word bundles are very common in the material (see Section 4). The present study focuses on bundles of three words, while the previous research tackles at least those bundles extending from three to eight words. Earlier studies on different university registers have centred on four-word strings, indicating that such bundles are more numerous than five-word bundles but offer more context for the investigation than three-word sequences do (Biber et al. 1999; Biber et al. 2003; Cortes 2004; Biber 2006 and Hyland 2008; see also Biber & Gray 2013). Studies on historical material, on the contrary, have included three-word bundles in their analyses (e.g. Culpeper & Kytö 2010: 106). Kopaczyk (2013a) covers bundles from three to eight words and the longer bundles are especially seen as manifestations of textual standardisation.

My study continues the research on three-word strings found in historical legal language. Three-word bundles were chosen as they offer enough data for analysis and their numbers are still manageable for a more detailed assessment. Further, they often appear as part of longer bundles (Biber et al. 1999: 992). The number of word sequences decreases noticeably in the material as the length of the bundle increases, as is shown in Table 1:

Table 1. Distribution of bundle lengths in the data

n-gram	<u>Early modern acts</u>		<u>2015 acts</u>	
	Types	Tokens	Types	Tokens
3-grams	1,187	31,940	452	6,606
4-grams	516	11,251	235	2,986
5-grams	206	4,222	134	1,533
6-grams	110	2,066	83	952
7-grams	68	1,137	63	762
8-grams	46	674	52	646
9-grams	30	391	47	577
10-grams	18	213	44	552
11-grams	11	128	42	531
12-grams	7	79	40	510
13-grams	2	23	38	489

In both groups of data, four-word bundles are at least about 50% less numerous than three-word strings. The longest bundles in the early modern acts are 13 words but in the contemporary material the longest occurrences extend to 45 words; these bundles are mainly different combinations of the enacting formula.

2.1 Bundles, legal language and genre conventions

Lexical bundles indicate that historical and present-day spoken and written texts often consist of prefabricated wording. In the registers of conversation and academic prose, for instance, approximately 30% and 21% of the wording respectively belong to repeated word strings (Biber et al. 1999: 995).¹ Recurrent communicative situations promote the use of formulaic wordings, and hence instead of choosing from individual words, speakers rely on prefabricated word

¹ Altenberg (1998: 102) presents a much higher estimate, stating that 80% of wordings are recurrent constructions in the *London-Lund Corpus* (see also e.g. Wray 2002: 28).

combinations that are retrieved as single chunks (e.g. Sinclair 1991: 110; Ellis 1996; Wray 2002: 9). Legal writing is generally consistent in its text type features because its authority is based on its prestigious form and because of the persistent purpose of acts to pose obligations on society. Hoey (2005: 129) further notes that established word strings often occur in specific parts of sentences or texts such as at the beginning of paragraphs. This chapter demonstrates that lexical bundles in acts are not only recurrent word combinations but also that many of them are firmly related to the textual patterning. The bundles show traces of textual organisation that extends beyond the level of words or sentences and can, therefore, be associated with a formulaic textual scheme. Cortes (2013) has earlier studied bundles in relation to Swales' (2004) moves in research article introductions.

Lexical bundles have been earlier analysed in both historical and present-day legal genres. Breeze (2013), for instance, focused on four contemporary subgenres of legal writing, including academic law, case law, legislation and documents. Legislation is found to be the most formulaic subgenre, containing the highest number of lexical bundles from four words to eight words (Breeze 2013: 233). Culpeper and Kytö (2010: 127) compared lexical bundles of Early Modern English dialogues to present-day trial proceedings and drama. Early modern trials rely more on narrative bundles and present-day trials on circumstantial and organisational word sequences, as trial proceedings have developed from narrative reports to cross-examination. Lexical bundles have been examined in historical written legislation as well. Kopaczyk (2013a) investigates bundles in the legal writing of the Scottish burghs between 1380 and 1560; the study focuses on formulaicity and standardisation and creates a formulaicity factor that distinguishes most formulaic bundles in the data by applying a token-to-file ratio and frequency counts on the bundles (see also Kopaczyk 2012 and 2013b).

2.2 Grammatical structure and the functions of lexical bundles

Lexical bundles have been mainly analysed in terms of their grammatical structures and functions. The study by Biber, Conrad and Cortes (2004) outlined a comprehensive taxonomy of bundles in university and classroom teaching and university textbooks. Their classification consists of three main syntactic classes and three functions. The grammatical types are verb phrase fragments, dependent clause fragments and noun and prepositional phrases; the functional bundles include stance expressions, discourse organisers and referential expressions (see also Biber et al. 2003). The three main grammatical categories are further divided into 17 subclasses, for instance, the category of noun and prepositional phrases includes separate classes for other noun phrase expressions (*a little bit more*) and prepositional phrase expressions (*of the things that*). Later studies applied this model to a number of data and genres (e.g. Cortes 2004). The grammatical categorisation defined in Biber et al. (2004) is used in my study as well because it fits the grammatical structures found in the data. However, I have added one extra category for binomial fragments that are often found in the material (see Section 6 for all the grammatical categories). Kopaczyk (2013a) similarly uses the categorisation in Biber et al. (2004) and adds a separate class for binomials.

More recent studies have established somewhat different functional classes to suit specific genres, but in line with Biber et al. (2004) the functions in these assessments are usually related to three basic aspects: references to extralinguistic factors, bundles centred on textual organisation and the relationship between the discourse participants (see e.g. Hyland 2008 and Culpeper & Kytö 2010: 110). The present study employs the categorisation used in Kopaczyk (2013a) since this framework reflects the bundles in the present data well. The functional categories in Biber et al. (2004), for instance, contain numerous stance expressions that are

important in academic writing but are rarely found in national legislation, which seeks to be more impersonal. Additionally, Breeze (2013) examines present-day legislation but merges the grammatical categories and functions, for example, references to extralinguistic concepts are grouped together with noun and prepositional phrases. In line with most of the previous studies, the taxonomy in Kopaczyk (2013a) distinguishes referential, textual and interactional bundles that refer to the Hallidayan framework with field, tenor and mode (see, e.g., Halliday 1978). Referential bundles thus refer to concepts outside the text and include subcategories, for instance, location, date and time and the authorities. Textual bundles organise the text and add cohesion such as primary cohesion (*of the said*) and narrative bundles (*and this is*). The referential and textual bundles often overlap since bundles such as *the said lands* refer anaphorically to earlier wordings in the texts and at the same time to an external reality. The interactional bundles include directive fragments (*and ordained that*) and deontic modality (*and if it*); these bundles stem from the speech act theory and they point to interaction between the participants (e.g. Searle 1976). I further added extra subcategories especially within the textual bundles to better reflect the present material. A class for quantities (*or more of*) was established under the referential bundles. An additional overlapping textual bundle category was added for word sequences including the word *such* (*of such lands*). Other new textual bundles are framing signals (*in respect of*) that limit the scope of the expression and resultative signals (*as a result*) (see Hyland 2008). Further, on the basis of the present-day acts, the textual category was amended to include categories for textual organisation and specific text references. The category of textual organisation mainly encompasses bundles that come from the beginning of marked sections, subsections or paragraphs ((A) *In the*), and word sequences on specific text references create textual cohesion by denoting the numbered parts in the text (*in subsection A*) (see Section 7 for the functional subcategories).

3. The Corpus of Early Modern English Statutes and Acts of Parliament from 2015

The historical material for this study is taken from *The Corpus of Early Modern English Statutes 1491–1707* (Lehto 2015a). The corpus includes both acts and proclamations, but only acts are scrutinised in the present chapter. The material totals 181,338 words, covering the years from 1491 to 1689 and containing altogether 153 acts. The enactments were taken from 18 Parliamentary collections printed after each Parliamentary session.² The documents encompass all the public acts enacted during the session, and the first 10,000 words from each collection were included.³ The corpus encompasses acts from 18% of the Parliamentary sessions that gathered during the Early Modern era, and 36% of the different Parliaments that were elected are represented in the corpus. Diachronically, the number of published acts increased during that period; however, the corpus contains an equal number of acts from both the sixteenth and seventeenth centuries since the priority was given to a balanced number of texts from both decades. In general, legal writing was in transition at the turn of the sixteenth century, as the language of the acts changed from French and Latin to English and as the texts started to be printed (see e.g. Elton 1983: 92–110). The acts were taken from the *Early English Books Online (EEBO)* and manually transcribed into text format.⁴

The Present-day English acts were retrieved from the online collections of the National Archives of Britain, which contains original and revised acts in text format. The first 16 public general acts

² Some of the documents contain acts from two or three Parliamentary sessions, increasing the number of covered Parliaments to 21.

³ Public acts concern the whole nation, while private acts apply only to certain areas or groups of people.

⁴ Most of the texts were not yet available in the *EEBO Text Creation Partnership* at the time of the corpus compilation, and hence the acts were transcribed from and proofread against the *EEBO* images.

from year 2015 were included in the analysis. The present-day data totals 55,371 words, but the word count for each act in the corpus was limited to 5,000.⁵ The word count for historical acts is higher since these documents cover more than 200 years of legal writing as opposed to the present-day acts that cover only a single year. The number of enacted public general acts in 2015 was 37, and thus the corpus encompasses material from about 43% of those acts.

4. Method of analysis

The bundles were retrieved using the n-grams tool in the corpus software AntConc (Anthony 2014). As the historical and present-day material are not comparable in size, different thresholds for frequency and distribution were set for both data. The analysis of the historical material covers bundles that appeared at least 10 times, which corresponds to 0.06 instances in 1,000 words. Further, the dispersion was adjusted to 5, that is to say, the word strings needed to occur at least in 5 different documents or about 28% of the data. In the present-day material, the frequency threshold was 4 occurrences, which corresponds to the normalised frequency of 0.07 in 1,000 words.⁶ The frequency is higher for the present-day acts; however, it was not sensible to lower the threshold in order to keep the distribution high enough. Hence, the minimum range was also 4. This threshold covers 25% of the contemporary acts, which is in line with the early modern material.⁷ The retrieved bundles were then categorised into grammatical and functional classes based on the studies by Biber et al. (2004) and Kopaczyk (2013a) (see Section 2.2).

The historical data was further analysed diachronically to see whether the grammatical categories or functions of the bundles changed in the course of the Early Modern era. For this assessment, the corpus was divided into two parts that roughly separated the sixteenth and seventeenth centuries. The subparts consisted of 9 documents: the sixteenth-century acts ranged from year 1491 to 1576 and the seventeenth century ones from 1593 to 1689. The frequency was set to 5 instances for both subparts, and dispersion was set to 4. The raw frequency of 5 is compatible with the frequency used for the early modern material as a whole, as it has a normalised frequency of 0.05 in 1,000 words in the sixteenth-century texts and 0.06 in the seventeenth century.⁸ The dispersion of 4 means that the bundles needed to occur in about 44% of the documents in each century; the bundles hence are rather formulaic, as they persist over numerous decades.

The historical texts show considerable variation in their orthography, although the spelling evolved towards standardisation in the latter half of the seventeenth century (see e.g. Görlach 1991 and Carney 1994). However, spelling variation in the corpus was fully normalised with the VARD (Variant Detector) software using automatic processing completed with manual standardisation (see e.g. Baron et al. 2009). It is worth noting that tabulation and numbering in the present-day data can affect the lexical bundles. As was shown in (1), many repeated wordings are elliptical due to the layout that can limit the number of occurrences of word strings. The cut-

⁵ As the corpus texts have been taken from the beginning of the acts, the corpus does not evenly represent all structural parts of present-day acts, such as the body of the text, amendments, schedules and definitions, which could have an effect on the distribution of the bundles.

⁶ The thresholds for frequency are somewhat higher in the present analysis than in earlier studies in which they vary from 10 to 40 instances per million words: the normalised frequency in the early modern data is 55.9 instances in a million words and 72.2 in the contemporary data (see Section 2).

⁷ The documents in the historical corpus were not divided into separate acts, since they varied in their word count, and doing so would cause difficulties in setting frequency and distribution cut-off points for a diachronic analysis. The earliest documents in the corpus contain several acts, but as the acts become diachronically much longer, the later texts include only 2 or 3 acts.

⁸ The frequency of the sixteenth-century texts is 55.9 in a million words and 56.0 in a million words in the seventeenth-century data.

off point of 4 is, however, rather low, and hence the elliptical structures likely do not extensively affect the results. In addition, the contemporary acts contain many numbers that refer, for instance, to amounts of money, but are also used to indicate textual organisation along with the letters. The numbers are, however, excluded from the bundles, as they could alter the results. Since numbering is also used on the textual level, the retrieved word combinations were analysed within the context of the text in order to trace those bundles related to the superstructure.

5. Distribution of lexical bundles in the historical and contemporary acts

The method of analysis retrieves altogether 1,639 bundle types for investigation. The early modern acts contain 1,187 bundles and 452 instances are found in the present-day acts. Further, the number of tokens in the data is almost 39,000 occurrences in total, as is indicated in Table 2:

Table 2. Number of bundle types and tokens in the data

3-grams	EModE acts	Present-day acts
Types	1,187	452
Tokens	31,940	6,606
% of tokens against the word count	17.9	11.9
% of types against the word count	0.7	0.8

The early modern and contemporary acts are not fully comparable, as different thresholds were used in their retrieval. However, the percentage of bundle types against the word count is rather similar in both sets of data, while the percentage of tokens is higher in the historical acts. The same lexical bundles are thus repeated more often in the early modern documents in which the bundles add up to 18% of the word count as opposed to 12% in the contemporary acts. The bundles can be less restated in the contemporary data, since these texts regularly use numbering and ellipses (see (1) above). The historical material additionally contains several acts on the same topic from different years, for instance, vagabonds are treated in various acts and result in bundles, such as *punishment of vagabonds*. The long diachronic coverage of the historical corpus does not seem to affect the number of bundles.

The most frequent bundles vary in the two eras, as is illustrated in the following Table 3:

Table 3. The fifteen most common bundles in the Early Modern and Present-day acts

<u>Early modern acts</u>			<u>Contemporary acts</u>		
Raw freq.	Range	Lexical bundle	Raw freq.	Range	Lexical bundle
646	18	of the said	171	14	in relation to
380	18	of the same	157	9	Secretary of State
275	18	person or persons	157	8	the Secretary of
253	18	by the authority	123	12	in respect of
226	18	enacted by the	102	13	the purposes of
192	14	and be it	87	11	section of the
182	18	in the said	87	10	under this section
161	17	of this act	71	13	is to be
		this present			
149	17	parliament	68	13	for the purposes
145	16	of this realm	65	11	to be treated

144	18	any of the	60	11	of that Act
123	18	according to the	60	11	the United Kingdom
123	18	be it enacted	59	10	be treated as
121	18	by the said	49	11	relation to a
118	18	of this present	46	7	in force at

The most common bundles found in the historical acts are anaphoric references to earlier statements in the documents (e.g. *of the said*, *of the same* and *in the said*). The binomial expression *person or persons* already indicates that the binomials are common within the bundles as well as bundles from the enacting formula stated in each act (e.g. *be it enacted*, *by the authority* and *this present parliament*). The lexical bundles are well established in the historical material, as most of the top bundles appear in all 18 documents, while in the contemporary data, the distribution is less extensive.⁹ The most frequent sequences in the contemporary acts are *in relation to* and *in respect of* that limit the extent of the law to certain cases, as in *This section does not apply in relation to a contract for the settlement*. Many of the top bundles further refer to certain numbered points in the acts such as *section of the* and *under this section* such as in *Part 5 of that act*.

The bundles are rather frequently encountered in the documents. This example comes from the Criminal Justice and Courts Act (2015), and the lexical bundles in the extract are italicized:¹⁰

- (2) 7 Electronic monitoring following release on licence etc
- (1) *Part 3 of the Criminal Justice and Court Services Act 2000 (dealing with offenders) is amended as follows.*
- (2) In section 62 (release on licence etc: conditions as to monitoring)—
- (a) *for subsection (2) substitute—*
- “(2) The conditions may include electronic monitoring conditions.
- (2A) An electronic monitoring condition imposed *under this section must* include a provision for making a person responsible for the monitoring.
- (2B) A person *may not be made* responsible for the monitoring unless *the person is* of a description specified in an order made *by the Secretary of State.*”, and
- (b) *after subsection (5) insert—*
- “(5A) In *this section* “electronic monitoring condition” means a condition requiring the person to submit to either or both of *the following—*
- (a) electronic monitoring of the person's compliance with another condition of release, and
- (b) electronic monitoring of the person's whereabouts (other than for the purpose of monitoring compliance with another condition of release).”

(2015 c. 2. Criminal Justice and Courts Act 2015, Part 1, Section 7)

This section of the act states the amendments to existing legislation and refers to numerous sections in the previous law using formulaic bundles including *for subsection x substitute* and *of the following*. In addition to lexical bundles, much of the wording in the extract is repeated, such

⁹ In Middle Scots legal texts, many formulaic sequences also range over all of the corpus texts from different geographical areas but longer bundles centre on the mid-eastern region, thus signalling a specific area for the spread of textual standardisation (Kopaczyk 2013a: 260).

¹⁰ The example contains italicized bundles that include more than three words, since the bundles overlap: *under this section must*, for example, encompasses the sequences *under this section* and *this section must* (for overlapping bundles, see Grabowski, this volume).

as *electronic monitoring of the person's compliance/whereabouts*; however, these structures do not appear within the bundles since they are not regularly used in the other acts.

6. Distribution of lexical bundles across different grammatical categories

A grammatical analysis of the early modern and present-day bundles indicates that most are noun phrase and prepositional phrase fragments, contributing to about 64% and 68% of the cases, respectively. The second most frequent grammatical types are verb phrase fragments, and these account for 20% of the historical acts and almost 25% of the acts in the modern documents. The third most common grammatical class is the dependent clause fragments followed by binomial fragments, as is shown in Table 4 below:

Table 4. Distribution of main grammatical categories of the lexical bundles

Grammatical categories	Early modern acts		Contemporary acts	
	%	Raw freq.	%	Raw freq.
NP and PrepP fragments	63.5%	754	67.5%	305
VP fragments	20.1%	238	24.8%	112
Dependent clause fragments	8.9%	105	5.3%	24
Binomial fragments	7.6%	90	2.4%	11

The proportion of dependent clause fragments and binomials is smaller in the contemporary acts. Dependent clauses are replaced by a systematic layout that uses numbering instead of repeating, for instance, the enacting formula and the subordinated clauses using *that* (*be it enacted that*). Further, as will be shown below, the percentage of dependent clause fragments begins to decline in the Early Modern period, while the frequency of binomials increases (see Section 8). In contemporary legal writing, however, these elements are superseded by other grammatical features. Noun and prepositional phrases have become more recurrent in present-day legislation, although they were already common in the historical acts, and account for more than 60% of the bundles. The percentage of verb phrases is also higher in the contemporary acts: the documents contain many amendments that include instructions for making changes to the existing law (e.g. *after subsection x insert*).

The grammatical categories were further analysed in more detail following the categorisation in Biber et al. (2004), as is illustrated in Table 5. The normalised frequencies for bundles are higher in the contemporary data, but comparison is problematic since different thresholds were used for retrieving the bundles.

Table 5. Detailed distribution of the grammatical categories of the lexical bundles

Grammatical type of lexical bundles	EModE acts		2015 acts	
	Raw freq.	f./1,000	Raw freq.	f./1,000
1. Verb phrase fragments	238	1.3	112	4.3
1a. (connector +) 1 st /2 nd person pronoun + VP fragment	0	0	0	0

1b. (connector +) 3 rd person pronoun + VP fragment	14	0.1	5	0.1
1c. Discourse marker + VP fragment	0	0	0	0
1d. VP with non-passive verb	119	0.7	70	1.3
1e. Verb phrase with passive verb	105	0.6	37	0.7
1f. yes/no question fragments	0	0	0	0
1g. WH-question fragments	0	0	0	0
2. Dependent clause fragments	105	0.6	24	0.4
2a. 1 st /2 nd person pronoun + dependent clause fragment	0	0	0	0
2b. WH-clause fragments	10	0.1	1	0.02
2c. <i>If</i> -clause fragments	13	0.1	5	0.1
2d. (verb/adjective+) to-clause fragment	23	0.1	16	0.3
2e. <i>That</i> -clause fragments	59	0.3	2	0.04
3. NP and PrepP fragments	754	4.2	305	5.5
3a. (connector +) Noun phrase with <i>of</i> -phrase fragment	180	1.0	85	1.5
3b. NP with other post-modifier fragment	17	0.1	30	0.5
3c. Other noun phrase expressions	272	1.5	57	1.0
3d. Prepositional phrase expressions	281	1.6	126	5.1
3e. Comparative expressions	4	0.02	7	0.1
4. Binomial fragments	90	0.5	11	0.2
Total	1,187	6.6	452	8.2

The frequency order of the lexical bundles differs somewhat between the historical and the present-day data. In both periods, within the most common grammatical types are prepositional phrase expressions, other noun phrase expressions and noun phrases with *of*-phrase fragments; additionally, active verb phrases are frequent in the contemporary material. Examples of prepositional phrase fragments include *of the said*, *by the queen*, *for the purposes*, *in relation to*, *to any person* and *by this section*. These bundles habitually function as anaphoric references to earlier parts of the texts. Additionally, the bundle *in this act* in the contemporary texts occurs at the beginning of definitions after a heading: the bundle and the definition are also placed on separate lines such as *In this act – “the appropriate authority means...”*. The bundle hence has a typical designated place in the paragraph in addition to being a recurrent lexical expression. Further, over one-third of the bundles in the historical texts begin with the preposition *of*, while only a few occurrences are encountered in the 2015 acts. The later acts favour other prepositions such as *in*, which often occurs as part of bundles that are referring to the text such as *in this part* and *in section a*. *Of*-constructions are, however, common as post-modifiers to noun phrases in the data. These findings reflect the general development of a post-modification to nouns and the prevalence of *of*-constructions. Prepositional phrases are the most common strategy of non-finite post-modification in the contemporary registers, and *of*-fragments are often used as they can express numerous functions (Biber et al. 1999: 634–635). The *of*-construction was already a typical noun post-modifier in formal written genres in the seventeenth century, but then other prepositions started to become more numerous and gained more varied meanings in later centuries and especially during the nineteenth century (Biber & Gray 2011: 241). Additionally, noun phrases with other post-modifier fragments are more important in contemporary legislation (*the person in*), signalling that the acts have developed a nominal character. The category of other

noun phrase expressions includes many references to the extralinguistic reality as well as anaphoric text references, for instance, *the said justices*, *any such person*, *King Henry the* and *this present parliament*. In the present-day material, there are additionally many references to the titles of other acts (e.g. *Criminal Justice Act*) and bundles of titles of amendments (*Annotations: Commencement Information*). Comparative expressions are generally rare in both sets of data.

Verb phrase fragments are numerous, as they are used to impose rights and duties. Both passive and active verb phrase fragments are employed in the historical data, while active verbs prevail in the contemporary texts where they are used to give instructions for the amendments. Active verb phrase fragments name the legal actors, for instance, in the bundle *the Treasury may* the legal actor treasury is given permission to amend the amounts of money mentioned in the act, as it is stated in *The Treasury may by order substitute a different amount for the amount for the time being specified in* (2015 act, c. 5). The verb phrase fragments with active verbs often consist of the verbs *shall*, *be* or *may* such as in the phrase *shall be lawful* and *this section may*. The subject is also often the act itself, such as *this section applies* and *subsection does not*. The passive verb phrase fragments again illustrate the importance of the amendments in the contemporary acts such as in the phrase *is amended as*. Grammatical structures that include the first or second person pronoun or discourse markers are absent in my data, although these elements are frequent in conversation (see Biber et al. 1999: 1003).

In the category of dependent clauses, the historical texts contain many *that*-clause fragments, while their contemporary counterparts favour *to*-clause fragments (see also Aarts, López-Couso & Méndez-Naya 2012: 878). The frequencies of *that*-clauses reflect the declining tendency of subordination and the increase of nominal elements in the written language (e.g. Biber & Gray 2011). *That*-clauses are further used often in the historical acts since they are part of the typical sentence structure and textual patterning, which are no longer found in the present-day acts. The orders are stated after the enacting formula in a subordinated *that*-clause, and each new subsection commences with a subordinated *that* clause, as can be seen below in (3):

- (3) Be it Enacted by the *Authority aforesaid*,
That from and after the Twenty
 ninth day of September next ensuing, no
 Brandies single or double in this Act
 mentioned shall be Imported from parts
 beyond the Seas in any Uessel or Cask,
 which shall not contain sixty Gallons at
 the least upon pain of forfeiting of the said
 Brandy or the value thereof, so to be Imported
 in small Uessels or Casks as aforesaid,
 whereof one half shall be to the
 Kings most Excellent Majesty, and the....

And be it Further Enacted and Ordained
 by the *Authority aforesaid*, *That*
 all and every the Officer and Officers
 who shall be concerned in the Levying,
 Collecting, and Receiving the Duties Arising
 by this Act, do keep a separate and
 distinct Account thereof, and Pay the...

(1685, James II, acts, ff. 50–51)

The bundle *authority aforesaid that* begins a new order in both paragraphs, and similar wordings extend beyond the example in the act. The bundles *be it enacted* and *and be it* also signal new subsections and are related to the superstructure of the acts. In the present-day data, these textual shifts are marked by numbering and layout. Further, the bundle *forasmuch as the* is found at the beginning of the historical acts: it indicates the beginning of a preamble in which the reasons for the act are stated. Similarly, the word string *provided always that* begins the new provisions (see (4) below). In the historical texts, the most common *to*-clause fragment is *to be made*, and it typically appears in the phrase *(grants) made or to be made*; yet these fragments also often define noun phrases in the early modern material (*leases to be made*). In the contemporary texts, *to*-clause fragments include the bundles *to be treated*, *to make provision* and *to be made*. The bundle *to be treated* appears as part of the phrase *to be treated as*, which orders how the law should be applied. *If*-clause fragments are rather rare in the data: in the historical texts they express a specific syntactic structure that expresses a hypothetical situation and an order for that situation, which follows the formula *if x – then y* as in the bundles *if any person – that then the* (see (4) below). The *if – then* structure is visible also in the contemporary acts, but the word *that* has been omitted from the main clause that follows. In the 2015 acts, *if*-clause fragments also illustrate textual organisation: the most frequent bundle is *if (a) the*, which shows the beginning of a new subsection, and further the bundle is divided into two separate lines, that is, *if* ends the previous line and *(a) the* is intended and placed on a separate line after the *if*.

Binomial fragments are especially frequent findings in Early Modern English acts, accounting for almost 8% of the 3-grams. In the historical data, the most frequent binomial fragments include the general noun *person* and all-inclusive attributive adjectives, that is, *person or persons*, *his or their* and *all and every*. The binomials also include fragments of multinomials, such as *city, town or*. As binomials and multinomials are frequent in the historical data, they generally refer to other concepts as well including authorities, legal actors, legal actions, land and other objects. The frequency of binomials further increases in the historical texts, although their role has diminished in the present-day acts (see Kopaczyk 2017: 174). The most common binomials in the 2015 acts are *on or after*, *England and Wales* and *advice and consent*, and these binomials hence refer to locations and contain formulaic wordings from the enacting formula. The binomials in both data sets are mainly noun phrases, while other grammatical categories are noticeably less frequent.

As is indicated by the grammatical categorisation, some of the lexical bundles begin with the connector *and* (and *or*). This trait is especially common in the historical acts, in which altogether 10% of the bundles commence with a connector, while in the contemporary acts that amount is 6%. Connectors are more prominent in the historical acts since they are used to signal new sections in the texts (e.g. *and that all persons, and be it enacted that*), while in the later documents the connectors have been replaced by tabulation and numbering. The decline of *and*-clauses further marks the development away from orality to a more literate writing style (see Lehto 2015b: 212). Consequently, in the historical data, *and* precedes mostly verb phrase fragments and other noun phrase expressions, while in the contemporary acts *and* is found at the beginning of more various types of noun phrases and prepositional phrases.

7. The functions of bundles in legal writing

Functional analysis reveals that most of the bundles in legal writing refer to concepts outside the actual text. In the historical acts, referential bundles account for 61% of the cases, while in contemporary legislation that proportion is lower, rising to 50%. When the bundles with textual overlap are taken into account, the percentages of the referential bundles rise to about 67% and 55% in the data, respectively. The second most common functional types are textual bundles:

these sequences are clearly more frequent in the contemporary acts (30.8%) than they are in the historical data (17.1%), as textual organisation and specific textual links become more important over time. Interactional bundles are rather common as well, rising to about 14% and 16%, as can be seen in Table 6:

Table 6. Distribution of the functions of the lexical bundles

Function	<u>EModE acts</u>		<u>2015 acts</u>	
	%	raw freq.	%	raw freq.
Referential bundles	60.8%	722	50.0%	226
Textual	17.1%	203	30.8%	139
Referential (textual overlaps)	5.9%	70	5.1%	23
Interactional bundles	16.2%	192	14.2%	64

Following the categorisation in Kopaczyk (2013a; see Section 2.2), the most common referential bundles in the early modern documents are actions, ambiguous cases, references to unnamed individuals and date and time, but other referential types are also found rather evenly. In the modern documents, actions are by far the most frequent functional type, but references to unnamed individuals and date and time are rather common findings as well (see Table 7):

Table 7. Detailed distribution of referential bundles

1. Referential bundles	<u>EModE acts</u>		<u>2015 acts</u>	
	Raw freq.	f./1,000	Raw freq.	Freq./1,000
a. date and time	87	0.5	20	0.4
b. collective reference to the authorities	59	0.3	14	0.3
c. collective reference to other community members	86	0.5	0	0
d. reference to named individuals	9	0.05	0	0
e. reference to unnamed individuals	91	0.5	26	0.5
f. action	122	0.7	114	2.1
g. location	58	0.3	9	0.2
h. object	81	0.5	12	0.2
i. ambiguous cases	118	0.7	17	0.3
j. quantity	11	0.06	14	0.3
Total	722	4.0	226	4.1

Referential bundles are common since the documents need to name various legal actors, actions, objects and other community members that the law affects, and the texts further situate these concepts within time and space. The bundles that refer to actions are most recurrent in the material: the historical texts contain an equal share of actions that are based on verb phrases (*appointed by the*) and those based on noun and prepositional phrases (*before the making*), while noun and prepositional phrases are more common in their modern counterparts. Actions that are expressed as noun phrase + *of* phrase fragments are numerous in both time eras (*the making of, the application of*), while other prepositional phrases are slightly more common in the contemporary data. References to unnamed individuals in both sets of the data mainly include the words *king, queen* and *person* (*to a person, by the queen*). The *Secretary of State* often

emerges in the contemporary data and *Justice of the Peace* appears in the historical material. Collective references to the authorities include officials and members of Parliament such as *this present parliament* and *and Commons in*. Many of these bundles stem from the enacting formula and are hence very similar in both data groups. Bundles that refer to objects in the historical acts encompass references to land and money and the acts as a piece of writing, such as *of this act* and *lands and tenements*. The category date and time includes bundles such as *date on which*, *the year of* and *the period of*. The function of quantity is especially common in the modern data and encompass examples such as *the amount of (tax)* and *of more than*. In the historical texts, references to named individuals include the king and the queen (*King Edward the*), and other community members also define groups of people and their role in society (*of the poor, offender and offenders, person and persons*), but these are absent in the contemporary data. Ambiguous cases are mainly bundles that do not show enough content words for functional analysis or have an unclear referent such as *every of them*. In general, most referential bundles are noun and prepositional fragments in both sets of data: in the 2015 acts, noun and prepositional bundles account for 39% of the cases and 51% in the historical documents.

The bundles that indicate textual functions display many differences between the historical and contemporary documents. In the early modern data, the most common textual bundles fall within the category of primary cohesion and intertextuality, and cohesive elements that overlap the referential function are frequent as well. In the contemporary texts, most of the textual bundles are references to specific parts in the same text or links to distinct sections in the earlier acts; specific text references are the second most frequent functional type in the contemporary acts after the bundles that describe actions. Further, intertextual links and framing signals are common in the contemporary documents, as can be seen in Table 8 below:

Table 8. Detailed distribution of textual bundles

2.Textual bundles	EModE acts		2015 acts	
	Raw freq.	f./1,000	Raw freq.	f./1,000
a. cohesion (primary)	95	0.5	10	0.2
b. cohesion (function overlaps)				
i. "said"	27	0.2	0	0
ii. "aforesaid"	6	0.03	0	0
iii. "same"	5	0.03	2	0.04
iv. "which/this/that/the very same"	14	0.08	19	0.3
v. "present/(this) same/the (one) in hand"	13	0.07	2	0.04
vi. "next"	4	0.02	0	0
vii. "such"	1	0.006	0	0
c. narration	14	0.08	4	0.07
d. intertextuality	81	0.5	28	0.2
e. clarification	3	0.02	1	0.02
f. framing signals	8	0.04	23	0.4
g. resultative signals	2	0.01	6	0.1
h. textual organisation	0	0	17	0.3
i. specific text reference	0	0	50	0.9
Total (textual bundles)	203	1.1	139	2.5
Total (referential with textual overlaps)	70	0.4	23	0.4

The cohesive bundles in the historical texts include wordings such as *and the said* and *of the aforesaid* that create general anaphoric links to earlier parts of those texts. Cohesive elements with an overlapping referential function are similarly frequent, including *our said sovereign* and *or said person*. The next example comes from an act from 1576 and it contains numerous general cohesive links, such as *the sayde person* and *in fourme aforesayd*:

- (4) Prouided alwayes neuerthelesse, that yf *thesayde person*
so committed, or to be committed, come before the next sessions of
the peace, or the next gaole deliuerie, to be holden for *thesayde*
Countie, or before their commyttyng, and doo fynde anye *suche*
suretie as is next before recited, to be bounde *infourmeaforesayd*,
for hym or her for one whole yeere: then *thesayde person* shal not
tarry in the gaole tyll the nexte sessions, or tyll the nexte gaole deliuerie.
(1576 acts, Elizabeth I, f. A6r)

The anaphoric link, *the said persons*, refers to vagabonds and their punishments, which is treated in the previous sections in the act before the provision, while at the same time the word string is a referential bundle signifying extralinguistic notions. These types of general links are less common in the modern acts, as the numbering allows for more detailed cohesive references to numbered parts of the texts, including *in paragraph (3a)* and *under section (255)B* (see (2) above). Intertextual relations are additionally more explicit in the contemporary acts with references to distinct paragraphs and sections in other acts, for example, *Article 3 of the Van Fuel Benefit Order 2014 (S.I. 2014/2896)* in which *order S I* appears as a lexical bundle, as the numbers are excluded. A further new function in the modern acts is the bundles that relate to textual organisation, for instance, *(a) in the* is a formulaic beginning of a new paragraph:

- (5) (4) The terms on which financial assistance may be provided include in particular—
(a) in the case of a grant or a loan, terms as to repayment;
(b) in the case of a guarantee, terms as to reimbursement.
(Acts of Parliament 2015, c. 7, Infrastructure act, Part I, 17)

The bundle *(a) in the* and *(b) in the* mark the beginnings of new paragraphs with a repeated beginning *in the case of a*. As tabulation and numbering developed, they enabled more specific textual cohesion and intertextual links within a single text and between other acts. The historical texts, however, already use the layout and visual means to signal new sections: the early modern acts often use paragraph changes or paragraph marks before new subsections, and the beginnings of new subsections are further marked by chaining clauses that begin with the conjunction *and* that functions similarly to the *(a) in the case* beginning in the above example (see also (3) above; see further Carroll et al. 2013; Lehto 2015b). Within the textual bundles, framing signals are important in the later acts, as these bundles set limitations on the scope of the law (e.g. *in relation to* and *in respect of* as in *applies only in relation to a person sentenced for an offence*). Another strategy in both eras is to use relative clauses to limit the extent of general nouns such as *the person who has received a payment*. The textual bundles in the historical data are most often created using prepositional phrase fragments and other noun phrase expressions. In the 2015 acts, prepositional phrases, noun phrases with *of*-phrase fragments and other noun phrase expressions prevail.

Interactional bundles in the acts indicate the imposing of orders to less powerful addressees. Modal bundles expressing deontic modality and directive bundles are the most often encountered types in the data, although modal bundles are more prominent in the historical acts (see Table 9):

Table 9. Detailed distribution of interactional bundles

3.Interactional bundles	EModE acts		2015 acts	
	Raw freq.	f./1,000	Raw freq.	f./1,000
a. directive bundles	54	0.3	30	0.5
b. representative bundles	7	0.04	0	0
c. declarative bundles	0	0	0	0
d. commissive bundles	0	0	0	0
e. modal bundles	131	0.7	34	0.6
Total	192	1.1	64	1.2

Some examples of the modal bundles are *act shall be*, *it may be* and *if*-clauses that impose duties under certain conditions, such as *if any person* (i.e. *if x – then y*, see (4) above). The verb *shall* is typically used in the early modern acts, while in the later acts *shall* is used only twice, and the verb *may* is favoured. Directive bundles include examples, such as *be it enacted*. Directions are hence expressed through indirect speech acts. In the contemporary acts, many of the directive bundles again reflect the giving of instructions on how to amend existing legislation such as *after subsection insert*. The directive bundles in the historical texts are typically formed using passive verbs and modals by active verbs. In the contemporary acts, modal bundles are similarly communicated by active verbs, but directive bundles show an equal share of both active and passive bundles, as the instructions for amendments are given in the active voice.

8. Diachronic development of lexical bundles in the early modern acts

The data from the Early Modern English period was examined diachronically, and the analysis retrieved 1,089 bundle types from the sixteenth-century texts and 986 occurrences from the seventeenth century. The frequency of both tokens and types against the word count is somewhat higher in the sixteenth-century documents, denoting that the earlier acts are more repetitive in their wording (see Table 10).

Table 10. Distribution of bundle types and tokens in the sixteenth and seventeenth centuries

3-grams	16 th century-acts	17 th century-acts
Types	1,089	986
Tokens	17,010	15,670
% of tokens against the word count	19.0	17.5
% of types against the word count	1.2	1.1

The top bundles in these two centuries are very similar if compared to the material as a whole. The most frequent bundles in the sixteenth century are *of the said*, *of the same*, *person or persons* and *by the authority*. In the seventeenth century, the top bundles are *of the said*, *of this act*, *of the same* and *person or persons*. In general, only about 17% of the sequences are not found among the bundles of the whole material. Even these bundles are very similar to the existing ones and many are new combinations of the same word strings.

The frequency order of the grammatical categories remains similar throughout the sixteenth and seventeenth centuries, but the proportions change to some extent. Noun and prepositional

fragments are the most common grammatical type found in both periods, followed by verb phrases, dependent clauses and binomials (see Table 11).

Table 11. Grammatical distribution of bundles in the sixteenth and seventeenth centuries

Grammatical type	16 th -century acts		17 th -century acts	
	%	Raw freq.	%	Raw freq.
NP and PrepP	63.7%	694	58.6%	578
Verb phrases	18.5%	201	23.0%	227
Dependent clauses	11.0%	120	9.8%	97
Binomial fragments	6.8%	74	8.5%	84

The percentages of noun and prepositional phrases and dependent clauses decline in the two centuries, and verb phrases and binomials become more frequent. The increase of binomials denotes that the texts emphasise all-inclusiveness by using formulaic word pairs by combining, for instance, synonyms or merely singular and plural forms. Most of the binomials are noun phrases but their proportion declines from 32% to 17%. This development contrasts with the evolvement of binomials in general in the acts, since the frequencies of especially the binomials that are noun phrases increase in early modern legal writing (Lehto 2017). The binomials that are nouns become more specific: the later documents refer to more distinct phenomena by word pairs, and these combinations do not occur as part of the 3-grams. Lexical bundles that are verb phrases additionally increase and demonstrate that formulaic wordings are used to impose orders to the citizens and refer to actions: these bundles are more general and include, for instance, the word string *persons shall be*.

At the same time, the proportion of formulaic noun and prepositional phrases decline: many of the noun and prepositional phrases are referential bundles, indicating that references to text-external reality by using noun and prepositional phrases become less important in the later acts. One reason for this change could be that the referents become more specific as the society became more diverse. The laws hence needed to cope and refer to more varied and complex concepts with nouns, while the indirect speech acts became more set. Dependent clause fragments further decrease in the material: *that* clauses and other dependent clauses are generally replaced by noun and prepositional phrase structures. This development is not encountered, however, in the proportions of the historical lexical bundles, denoting that many of these structures fail to become formulaic in the material or that the development takes place later (see Section 9).

The functions of the sixteenth and seventeenth-century bundles reflect the changes occurring in the grammatical types of the word strings. Referential bundles decrease and interactional bundles increase, as the referential bundles are typically expressed by noun and prepositional phrases and the interactional bundles are expressed by verb phrase fragments (see Table 12).

Table 12. Functions of the sixteenth and seventeenth-century bundles

Function	16 th -century acts		17 th -century acts	
	%	Raw freq.	%	Raw freq.
Referential	63.8%	695	60.0%	592
Textual	15.1%	164	15.1%	149
Referential with textual overlaps	6.5%	71	5.9%	58
Interactional	14.6%	159	19.0%	187

The percentages for the referential bundles with textual overlaps decrease slightly. The proportion of textual bundles in both centuries remains similar, as the acts continue to rely on general anaphoric links. The development towards more specific textual references used in contemporary acts does not yet take place in the Early Modern era.

9. Discussion and conclusion

The analysis encompassed a rather high number of bundles with 1,639 instances, and the early modern data was further analysed diachronically. Both Early Modern and Present-day English legal writing is based mainly on noun and prepositional constructions, as these add up to about 64% and 68% of the cases, respectively. Dependent clause fragments and binomials are less important in the contemporary legal texts. In the historical material, dependent clauses are repeated constructions since they express orders after the enacting formula in the following *that*-clause. In the contemporary texts, these sentence relations are expressed by the layout. Verb phrases are more recurrent in the present-day acts, and they are used especially in amendments to give instructions on the changes to the earlier version of the act. Many of the bundles in the historical texts further begin with the connector *and*, indicating that the sentence structure was based on addition: new subsections were signalled by the conjunction *and*, followed by a shortened version of the enacting formula (*and be it also enacted*). In the contemporary texts, these shifts are again displayed by layout. The decline of *and* at the beginning of the bundles further reflects a change towards literacy, since coordinated clauses are preferred in spoken genres and phrasal coordination is preferred in written registers (e.g. Biber et al. 1999: 81). There is still much continuation in legal writing, as the proportions of the main grammatical categories of the bundles remain rather constant in the data.

The most common functional category in legal texts is the referential bundles. This finding is expected since the acts give orders for abstract and concrete objects, actions and people that exist outside the legal texts. The importance of the textual level is, however, visible, as textual bundles are the second most common category, rising to 31% in the contemporary acts and to 17% in the historical material. Textual references become more detailed in the contemporary acts, since the bundles contain numbers and letters that refer to specific parts of the orders or form inter-textual links to earlier acts. In the historical data, the bundles create cohesion, for instance, with the more general words *said* and *aforsaid*. In the contemporary acts, specific references and also amendments to existing laws are more crucial, and formulaic phrasings emerge that announce changes to previous acts. Instead of enacting completely new legislation, many of the acts make amendments, as there is a chain of acts appearing on the same topic. Textual organisation is further shown in the bundles that contain numbers that emerge from the detailed structuring of the orders that developed in the nineteenth century: the numbers and letters indicate the beginning of a new subsection and often show an ellipsis. In the historical data, the textual organisation surfaces through the grammatical structure, including *that*-clause fragments and bundles such as *and be it enacted*; in the contemporary material, the textual organisation is visible in the functional categorisation through bundles that are titled as textual organisation.

The main distribution of the grammatical categories stays rather similar in the early modern era. However, the proportions of noun and prepositional phrases and dependent clauses are lower in the seventeenth century than in the previous century, and verb phrases and binomials increase in their proportions. The decline of dependent clauses can reflect a change toward a nominal style of writing, but the decrease of noun and prepositional phrase fragments is not in line with this

development. Written genres evolved toward nominal features at least from the eighteenth century onwards, when nominalisations and prepositional phrases as post-modifiers with *of* were common; further, in the nineteenth century, other prepositional phrases used as post-modifiers increased along with nouns used as nominal pre-modifiers (see Biber & Gray 2011). This change can be seen in the present-day acts in the data, as the proportion of noun and prepositional phrase fragments is higher than in the sixteenth or seventeenth centuries and noun phrases with other post-modifier fragments are more numerous. Further, dependent clauses declined, although verb phrase fragments are also common. The change toward this nominal writing style thus possibly took place after the Early Modern English era.

The historical legal texts are already distinguished from spoken legal language and reflect written conventions. The distribution of the grammatical and functional categories in my study is in line with Scots legal texts, in which referential bundles and noun and prepositional phrases prevail (see Kopaczyk 2013a: 162–176, 235). The findings contrast with the historical spoken language, as Culpeper and Kytö (2010: 118) state that noun and prepositional phrase fragments are rare, and trial proceedings prefer verb phrase fragments. The verb phrases in trials further include personal pronouns and they are typically found as part of questions. Verb phrase fragments are additionally found to be common in Present-day English spoken language (Biber et al. 2004: 380–382; see also Hyland 2008).

Lexical bundles in legal writing need to be analysed in light of the textual structure and layout, as the word strings also display fragments of textual organisation in addition to revealing formulaic word combinations. The bundles with *if*-clause fragments, for instance, not only become more specific but also change in their layout. In the contemporary material, these bundles often contain ordered

lists that mark new paragraphs of an act and are placed on different lines (e.g. *if – (a) the*). The influence of the textual structure is, nevertheless, also visible in the historical data where bundles beginning with *and* often mark new subsections of the acts and the bundle *forasmuch as the*, for instance, begins preambles. These sections are often signalled through the layout by paragraph marks, gaps in the running text and paragraph changes, although tabulation with numbering was not yet in use at the time. The focus of the contemporary acts is often on the text itself, as textual links and amendments are common.

The diachronic changes recognised in this study stem both from general linguistic developments that were taking place in the Early Modern and Present-day English eras and the sociohistorical changes that were affecting the legal system. The textual structure mostly reflects new conventions in the legal tradition: the special layout with textual links emerged, as the legal field became more dependent on written documents and needed detailed referencing and intertextual links. At the same time, the acts developed toward the nominal writing conventions in line with other written genres, that is, written language differentiated from spoken genres, illustrated, for instance, by the increase of prepositional phrases in the acts. Some of the diachronic changes originated both from general linguistic changes and novel practices in the writing conventions of the acts. The decline of dependent clauses with *that*, for instance, follows a trend found in other written genres, but it also reflects changes made to the layout that lessened the need to repeat the enacting formula. The acts hence illustrate many formulaic elements and continuity in legal writing, but the lexical bundles also reveal numerous diachronic differences caused by sociohistorical changes and developments in written genres.

Corpora and tools

- Anthony, Laurence. 2014. AntConc 3.4.3. Corpus software. Tokyo: Waseda University. <<http://www.laurenceanthony.net/>>
- Early English Books Online (EEBO). <<http://eebo.chadwyck.com>>
- Early English Books Online – Text Creation Partnership (EEBO-TCP). <<http://quod.lib.umich.edu/e/eebogroup/>>
- The National Archives. Records of Public General Acts. <legislation.gov.uk>

References

- Aarts, Bas, López-Couso, María José & Méndez-Naya, Belén. 2012. Late Modern English: Syntax. In *English Historical Linguistics*, Vol. I, Alexander Bergs & Laurel J. Brinton (eds), 869–887. Berlin: Mouton de Gruyter.
- Altenberg, Bengt. 1998. On the phraseology of spoken English: The evidence of current word-combinations. In *Phraseology: Theory, Analysis and Applications*, Anthony P. Cowie (ed.), 100–122. Oxford: OUP.
- Baron, Alistair, Rayson, Paul & Archer, Dawn. 2009. Automatic standardization of spelling for historical text mining. In *The Proceedings of Digital Humanities 2009*, 309–312. College Park MD: University of Maryland.
- Biber, Douglas. 2006. *University Language: A Corpus-based Study of Spoken and Written Registers* [Studies in Corpus Linguistics 23]. Amsterdam: John Benjamins. doi: 10.1075/scl.23
- Biber, Douglas, Conrad, Susan & Cortes, Viviana. 2004. If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics* 25(3): 371–405. doi:10.1093/applin/25.3.371
- Biber, Douglas, Conrad, Susan & Cortes, Viviana. 2003. Lexical bundles in speech and writing: An initial taxonomy. In *Corpus Linguistics by the Lune: A Festschrift for Geoffrey Leech* [Łódź Studies in Language 8], Andrew Wilson, Paul Rayson & Tony McEnery (eds), 71–105. Frankfurt: Peter Lang.
- Biber, Douglas & Gray, Bethany. 2013. Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics* 18(1): 109–135. doi: 10.1075/ijcl.18.1.08gra
- Biber, Douglas & Gray, Bethany. 2011. Grammar emerging in the noun phrase: The influence of written language use. *English Language and Linguistics* 15(2): 223–250. doi: 10.1017/S1360674311000025
- Biber, Douglas, Johanson, Stig, Leech, Geoffrey, Conrad, Susan & Finegan, Edward. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Breeze, Ruth. 2013. Lexical bundles across four legal genres. *International Journal of Corpus Linguistics* 18(2): 229–253. doi: 10.1075/ijcl.18.2.03bre
- Carney, Edward. 1994. *A Survey of English Spelling*. London: Routledge. doi: 10.4324/9780203199916
- Carroll, Ruth, Peikola, Matti, Salmi, Hanna, Varila, Mari-Liisa, Skaffari, Janne & Hiltunen, Risto. 2013. Pragmatics on the page. *European Journal of English Studies* 17(1): 54–71. doi: 10.1080/13825577.2013.755006
- Cortes, Viviana. 2013. *The purpose of this study is to*: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes* 12(1): 33–43. doi: 10.1016/j.jeap.2012.11.002
- Cortes, Viviana. 2004. Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes* 23(4): 397–423. doi: 10.1016/j.esp.2003.12.001

- Culpeper, Jonathan & Kytö, Merja. 2010. *Early Modern English Dialogues: Spoken Interaction as Writing*. Cambridge: CUP.
- Danet, Brenda. 1980. Language in the legal process. *Law and Society Review* 14(3): 445–564. doi:10.2307/3053192
- van Dijk, Teun A. 1985. Structures of news in the press. In *Discourse and Communication: New Approaches to the Analysis of Mass Media Discourse and Communication*, Teun A. van Dijk (ed.), 69–93. Berlin: Walter de Gruyter. doi: 10.1515/9783110852141.69
- Ellis, Nick C. 1996. Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition* 18(1): 91–126. doi: 10.1017/S0272263100014698
- Elton, Geoffrey Rudolph. 1983. *Studies in Tudor and Stuart Politics and Government, Vol. 3: Papers and Reviews 1973–1981*. Cambridge: CUP. doi: 10.1017/CBO9780511560514
- Görlach, Manfred. 1991. *Introduction to Early Modern English*. Cambridge: CUP. doi: 10.1017/CBO9781139166010
- Halliday, Michael A. K. 1978. *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.
- Hoey, Michael. 2005. *Lexical Priming: A New Theory of Words and Language*. New York NY: Routledge. doi:10.4324/9780203327630
- Hyland, Ken. 2008. *As can be seen*: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27: 4–21. doi: 10.1016/j.esp.2007.06.001
- Kopaczyk, Joanna. 2013a. *The Legal Language of Scottish Burghs: Standardization and Lexical Bundles 1380–1560*. Oxford: OUP. doi: 10.1093/acprof:oso/9780199945153.001.0001
- Kopaczyk, Joanna. 2013b. Formulaic discourse across Early Modern English medical genres: Investigating shared lexical bundles. In *Meaning in the History of English: Words and Texts in Context* [Studies in Language Companion Series 148], Andreas H. Jucker, Daniela Landert, Annina Seiler & Nicole Studer-Joho (eds), 257–299. Amsterdam: John Benjamins. doi:10.1075/slcs.148.12kop
- Kopaczyk, Joanna. 2012. Long lexical bundles and standardisation in historical legal texts. *Studia Anglica Posnaniensia* 47(2–3): 3–26.
- Kopaczyk, Joanna. 2017. Terms and conditions: A comparative study of noun binomials in UK and Scottish legislation. In *Phraseology in Legal and Institutional Settings: A Corpus-based Perspective*, Stanisław Goźdz-Roszkowski & Gianluca Pontrandolfo (eds), 160–185. London and New York: Routledge.
- Lehto, Anu. 2015a. Corpus of Early Modern English Statutes (1491–1707). Unpublished corpus. University of Helsinki.
- Lehto, Anu. 2015b. *The Genre of Early Modern English Statutes: Complexity in Historical Legal Writing* [Mémoires de la Société Néophilologique de Helsinki 97]. Helsinki: Modern Language Society.
- Lehto, Anu. 2017. Binomials and multinomials in Early Modern English parliamentary acts. In *Binomials in the History of English: Fixed and Flexible*, Joanna Kopaczyk & Hans Sauer (eds), 241–260. Cambridge: CUP.
- McEnery, Tony, Xiao, Richard & Tono, Yukio. 2006. *Corpus-based Language Studies: An Advanced Resource Book*. New York NY: Routledge.
- Meyer, Charles F. 2015. Corpus-based and corpus-driven approaches to linguistic analysis: One and the same? In *Developments in English: Expanding Electronic Evidence*, Irma Taavitsainen, Merja Kytö, Claudia Claridge & Jeremy Smith (eds), 14–28. Cambridge: CUP.
- Moon, Rosamund. 1998. *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon Press.
- Searle, John R. 1976. Classification of illocutionary acts. *Language in Society* 5: 1–23. doi: 10.1017/S0047404500006837

- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: OUP.
- Swales, John. 2004. *Research Genres: Exploration and Applications*. Cambridge: CUP. doi: 10.1017/CBO9781139524827
- Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work* [Studies in Corpus Linguistics 6]. Amsterdam: John Benjamins. doi: 10.1075/scl.6
- Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge: CUP. doi: 10.1017/CBO9780511519772